

Uniqueness and Monge solutions in the multi-marginal optimal transportation problem*

BRENDAN PASS[†]

August 27, 2010

Abstract

We study a multi-marginal optimal transportation problem. Under certain conditions on the cost function and the first marginal, we prove that the solution to the relaxed, Kantorovich version of the problem induces a solution to the Monge problem and that the solutions to both problems are unique.

1 Introduction

Given two Borel probability measures μ_1 and μ_2 on spaces M_1 and M_2 respectively and a cost function $c : M_1 \times M_2 \rightarrow \mathbb{R}$, Monge's optimal transportation problem asks how to most efficiently map μ_1 onto μ_2 , where efficiency is measured relative to c . We say that a function $G : M_1 \rightarrow M_2$ pushes μ_1 forward to μ_2 and write $G_{\#}\mu_1 = \mu_2$ if $\mu_1(G^{-1}(B)) = \mu_2(B)$ for all measurable $B \subseteq M_2$. Monge's problem is then to minimize $\int_{M_1} c(x_1, G(x_1)) d\mu_1$ among all G that push μ_1 forward to μ_2 . Due to both its deep connections to other areas of mathematics and its applicability in other fields, optimal

*The author was supported in part by an NSERC postgraduate scholarship. This work was completed in partial fulfillment of the requirements for a doctoral degree in mathematics at the University of Toronto.

[†]Department of Mathematics, University of Toronto, Toronto, Ontario, Canada, M5S 2E3 bpass@math.utoronto.ca.

transportation has grown into a thriving field of research over the past 20 years; we refer the interested reader to the books by Villani for references [21][22].

In this paper, we are interested in a multi-marginal version of the preceding problem. Given Borel probability measures μ_i on n -dimensional, smooth manifolds M_i , for $i = 1, 2, \dots, m$, and a *cost* function $c : M_1 \times M_2 \times \dots \times M_m \rightarrow \mathbb{R}$, the multi-marginal version of Monge's optimal transportation problem is to minimize:

$$C(G_2, G_3, \dots, G_m) := \int_{M_1} c(x_1, G_2(x_1), G_3(x_1), \dots, G_m(x_1)) d\mu_1 \quad (\mathbf{M})$$

among all $(m - 1)$ -tuples of measurable maps (G_2, G_3, \dots, G_m) , where $G_i : M_1 \rightarrow M_i$ pushes μ_1 forward to μ_i for all $i = 2, 3, \dots, m$. The Kantorovich formulation of the problem is to minimize:

$$C(\mu) := \int_{M_1 \times M_2 \times \dots \times M_m} c(x_1, x_2, x_3, \dots, x_m) d\mu \quad (\mathbf{K})$$

among all positive Borel measures μ on $M_1 \times M_2 \times \dots \times M_m$ such that the canonical projection

$$\pi_i : M_1 \times M_2 \times \dots \times M_m \rightarrow M_i$$

pushes μ forward to μ_i for all i . For any $(m - 1)$ -tuple (G_2, G_3, \dots, G_m) such that $G_{i\#}\mu_1 = \mu_i$ for all $i = 2, 3, \dots, m$, we can define the measure $\mu = (Id, G_2, G_3, \dots, G_m)_{\#}\mu_1$ on $M_1 \times M_2 \times \dots \times M_m$, where $Id : M_1 \rightarrow M_1$ is the identity map. Then μ projects to μ_i for all i and $C(G_2, G_3, \dots, G_m) = C(\mu)$; therefore, **K** can be interpreted as a relaxed version of **M**. Roughly speaking, the difference between the two formulations is that in **M** almost every point $x_1 \in M_1$ is coupled with exactly one point $x_i \in M_i$ for each $i = 2, 3, \dots, m$, whereas in **K** an element of mass at x_i is allowed to be *split* between two or more target points in M_i for $i = 2, 3, \dots, m$.

Assuming that c is continuous, it is not hard to show that a solution to **K** exists. When $m = 2$, under a regularity condition on μ_1 and a twist condition on c , which we will define in the next section, one can show that this solution is concentrated on the graph of a function over x_1 [14][7][1][8][2]. It is then straightforward to show that this function solves **M** and to establish uniqueness results for both **M** and **K**. When $m \geq 3$, however, existence and uniqueness in **M** as well as uniqueness in **K** are still

largely open. In their seminal paper, Gangbo and Świąch [9] used a duality theorem of Kellerer [11] to resolve these questions for the quadratic cost function, $c(x_1, x_2, x_3, \dots, x_m) = -|\sum_{i=1}^m x_i|^2$ on $\mathbb{R}^n \times \mathbb{R}^n \times \dots \times \mathbb{R}^n$, extending partial results for the same cost by Olkin and Rachev [17], Knott and Smith [13] and Rüschendorf and Uckelmann [19]. Gangbo and Świąch's theorem was then reproved using a different argument by Rüschendorf and Uckelmann [20] and generalized by Heinich [10] to cost functions of the form $c(x_1, x_2, x_3, \dots, x_m) = h(\sum_{i=1}^m x_i)$ where $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is strictly concave. Carlier extended these results to a still wider class of cost functions, but only in the case when all the domains M_i are one dimensional [3]. A result with a somewhat different flavour was obtained by Carlier and Nazaret [5]. Assuming $n = m$ and $M_i = \mathbb{R}^n$ for all i , their objective was to maximize a convex function of the determinant of the matrix whose columns are the vectors x_i ; for this cost function, they demonstrate that the maximizer may not be concentrated on the graph of a function over x_1 and may not be unique. The aim of the present article is to identify general conditions on c under which both **K** and **M** admit unique solutions.

With one exception, the conditions we impose will look similar to standard conditions which arise when studying the two marginal problem. Our lone novel hypothesis is that a certain covariant 2-tensor on the product space $M_2 \times M_2 \times \dots \times M_{m-1}$ should be negative definite. Tensors have recently become very relevant to optimal transportation, owing to their essential role in the study of the regularity of optimal maps. Ma, Trudinger and Wang [15] showed that a certain tensorial condition dictates the regularity of the solution to **M** in the two marginal case, and Kim and McCann [12] reinterpreted this condition, relating it to the sectional curvature of a certain pseudo-metric. In [18], the present author showed that the dimension of the support of the optimizers for multi-marginal problems is related to a family of symmetric, bi-linear forms. One surprising consequence of that work is a class of counterexamples demonstrating that the obvious generalization of the twist condition to the multi-marginal setting is sufficient neither to guarantee uniqueness of minimizers for **K** nor to ensure that the solution to **K** induces a solution to **M**. In these examples, the solutions were concentrated on submanifolds of dimension greater than n ; motivated in part by this observation, [18] identified local conditions on c under which the support of the optimal measure is at most n -dimensional. Our condition here is a little different. Whereas the question about the dimension of the support of a solution μ to **K** is purely local, showing that μ gives rise to a solution to **M**

is a global issue: for almost all $x_1 \in M_1$ we must show that there is exactly one $(x_2, x_3, \dots, x_m) \in M_2 \times M_3 \times \dots, M_m$ which get coupled to x_1 by μ . Our tensor here is designed to capture this global nature of the problem.

In the next section we recall relevant concepts from the theory of optimal transportation and formulate the conditions we will need. In section 3 we state and prove our main result and in the fourth section we exhibit several examples of cost functions which satisfy the criteria of our main theorem.

Acknowledgment: It is a pleasure to thank my advisor, Robert McCann, for his support and for many useful discussions throughout the duration of this work.

2 Preliminaries and definitions

We will assume that each M_i can be smoothly embedded in some larger manifold in which its closure $\overline{M_i}$ is compact and that the cost $c \in C^2(\overline{M_1} \times \overline{M_2} \times \dots \times \overline{M_m})$. In addition, we will assume that M_i is a Riemannian manifold for $i = 2, 3, \dots, m-1$ and that any two points can be joined by a smooth, length minimizing geodesic¹, although no such assumptions will be needed on M_1 or M_m . The requirement of a Riemannian structure is related to the global nature of \mathbf{M} that we alluded to in the introduction; a Riemannian metric gives us a natural way to connect any pair of points, namely geodesics.

We will denote by $D_{x_i}c(x_1, x_2, \dots, x_m)$ the differential of c with respect to x_i . For $i \neq j$, the bi-linear form $D^2_{x_i x_j}c(x_1, x_2, \dots, x_m)$ on $T_{x_i}M_i \times T_{x_j}M_j$ was introduced in [12]; in local coordinates, it is defined by

$$D^2_{x_i x_j}c\left(\frac{\partial}{\partial x_i^{\alpha_i}}, \frac{\partial}{\partial x_j^{\alpha_j}}\right) = \frac{\partial^2 c}{\partial x_i^{\alpha_i} \partial x_j^{\alpha_j}}.$$

As M_i is Riemannian for $i = 2, \dots, m-1$, Hessians or unmixed, second order partial derivatives with respect to these coordinates make sense and we will denote them by $Hess_{x_i}c(x_1, x_2, \dots, x_m)$; note, however, that no Riemannian structure is necessary to ensure the tensoriality of the mixed second order partials $D^2_{x_i x_j}c(x_1, x_2, \dots, x_m)$, as was observed in [12].

Given a Borel probability measure μ on a topological space M , the support of μ , which we will denote by $spt(\mu)$, is defined to be the smallest closed subset of M such that $\mu(spt(\mu)) = 1$.

¹Note that we do *not* assume M_i is complete, however, as we do not wish to exclude, for example, bounded, convex domains in \mathbb{R}^n .

The dual problem to **K** is to maximize

$$\sum_{i=1}^m \int_{M_i} u_i(x_i) d\mu_i \quad (\mathbf{D})$$

among all m -tuples (u_1, u_2, \dots, u_m) of functions $u_i \in L^1(\mu_i)$ for which $\sum_{i=1}^m u_i(x_i) \leq c(x_1, \dots, x_m)$ for all $(x_1, \dots, x_m) \in M_1 \times M_2 \times \dots \times M_m$.

There is a special class of functions satisfying the constraint in **D** that will be of particular interest to us:

Definition 2.1. *We say that an m -tuple of functions (u_1, u_2, \dots, u_m) is c -conjugate if for all i*

$$u_i(x_i) = \inf_{\substack{x_j \in M_j \\ j \neq i}} \left(c(x_1, x_2, \dots, x_m) - \sum_{j \neq i} u_j(x_j) \right)$$

Whenever (u_1, u_2, \dots, u_m) is c -conjugate, the u_i are semi-concave and hence have super differentials $\overline{\partial}u_i(x_i)$ at each point $x_i \in M_i$. By compactness, for each $x_i \in M_i$ we can find $x_j \in \overline{M_j}$ for all $j \neq i$ such that $u_i(x_i) = c(x_1, x_2, \dots, x_m) - \sum_{j \neq i} u_j(x_j)$; furthermore, as long as $|u_i(x_i)| < \infty$ for at least one x_i , u_i is locally Lipschitz [16].

The following theorem makes explicit the link between the Kantorovich problem and its dual.

Theorem 2.1. *There exists a solution μ to the Kantorovich problem and a c -conjugate solution (u_1, u_2, \dots, u_m) to its dual. Furthermore, the maximum value in **D** coincides with the minimum value in **K**. Finally, for any solution μ to **K**, any c -conjugate solution (u_1, u_2, \dots, u_m) to **D** and any $(x_1, \dots, x_m) \in \text{spt}(\mu)$ we have $\sum_{i=1}^m u_i(x_i) = c(x_1, \dots, x_m)$.*

This result is well known in the two marginal case; for $m \geq 3$, the existence of solutions to **K** and **D** as well as the equality of their extremal values was proved in [11]. The remaining conclusions were proved for a special cost by Gangbo and Świąch [9] and for a general, continuous cost when each $M_i = \mathbb{R}^n$ by Carlier and Nazaret [5]. The same proof applies for more general spaces M_i ; we reproduce it below in the interest of completeness.

Proof. As mentioned above, a proof of the existence of solutions μ to **K** and (v_1, v_2, \dots, v_m) to **D** as well as the equality:

$$\sum_{i=1}^m \int_{M_i} v_i(x_i) d\mu_i = \int_{M_1 \times M_2 \times \dots \times M_m} c(x_1, x_2, x_3, \dots, x_m) d\mu \quad (1)$$

can be found in [11]. We use a convexification trick, also found in [9] and [5], to build a c -conjugate solution to \mathbf{D} .

Define

$$u_1(x_1) = \inf_{\substack{x_j \in M_j \\ j \geq 2}} \left(c(x_1, x_2, \dots, x_m) - \sum_{j=2}^m v_j(x_j) \right)$$

and u_i inductively by

$$u_i(x_i) = \inf_{\substack{x_j \in M_j \\ j \neq i}} \left(c(x_1, x_2, \dots, x_m) - \sum_{j=1}^{i-1} u_j(x_j) - \sum_{j=i+1}^m v_j(x_j) \right)$$

As

$$u_m(x_m) = \inf_{\substack{x_j \in M_j \\ j \neq m}} \left(c(x_1, x_2, \dots, x_m) - \sum_{j=1}^{m-1} u_j(x_j) \right),$$

we immediately obtain

$$u_i(x_i) \leq \inf_{\substack{x_j \in M_j \\ j \neq i}} \left(c(x_1, x_2, \dots, x_m) - \sum_{j \neq i} u_j(x_j) \right). \quad (2)$$

The definition of u_{i-1} implies that for all (x_1, x_2, \dots, x_m)

$$v_i(x_i) \leq c(x_1, x_2, \dots, x_m) - \sum_{j=1}^{i-1} u_j(x_j) - \sum_{j=i+1}^m v_j(x_j)$$

Therefore, $v_i(x_i) \leq u_i(x_i)$. It then follows that

$$\begin{aligned} u_i(x_i) &= \inf_{\substack{x_j \in M_j \\ j \neq i}} \left(c(x_1, x_2, \dots, x_m) - \sum_{j=1}^{i-1} u_j(x_j) - \sum_{j=i+1}^m v_j(x_j) \right) \\ &\geq \inf_{\substack{x_j \in M_j \\ j \neq i}} \left(c(x_1, x_2, \dots, x_m) - \sum_{j \neq i} u_j(x_j) \right), \end{aligned}$$

which, together with (2), implies that (u_1, u_2, \dots, u_m) is c -conjugate. Now, we

have

$$\begin{aligned}
\sum_{i=1}^m \int_{M_i} v_i(x_i) d\mu_i &\leq \sum_{i=1}^m \int_{M_i} u_i(x_i) d\mu_i \\
&= \sum_{i=1}^m \int_{M_1 \times M_2 \dots \times M_m} u_i(x_i) d\mu \\
&\leq \int_{M_1 \times M_2 \dots \times M_m} c(x_1, x_2, x_3, \dots, x_m) d\mu
\end{aligned}$$

and so by (1) we must have

$$\sum_{i=1}^m \int_{M_i} u_i(x_i) d\mu_i = \sum_{i=1}^m \int_{M_1 \times M_2 \dots \times M_m} u_i(x_i) d\mu = \int_{M_1 \times M_2 \dots \times M_m} c(x_1, x_2, x_3, \dots, x_m) d\mu$$

But because $\sum_{i=1}^m u_i(x_i) \leq c(x_1, x_2, x_3, \dots, x_m)$, we must have equality μ almost everywhere. Continuity then implies equality holds on $\text{spt}(\mu)$. \square

As a corollary to the duality theorem, we now prove a uniqueness result for the solution to **D**. When $m = 2$, this result, under the weak conditions on c stated below, is due to Chiappori, McCann and Nesheim [6]; for certain special, multi-marginal costs, it was proven by Gangbo and Świąch [9] and Carlier and Nazaret [5]. Although this result is tangential to the main goals of this article, we prove it here to emphasize that, whereas uniqueness in **K** requires certain structure conditions on the cost, uniqueness in **D** depends only on the differentiability of c .

Corollary 2.1. *Suppose the domains M_i are all connected, that c is continuously differentiable and that each μ_i is absolutely continuous with respect to local coordinates with a strictly positive density. If (v_1, v_2, \dots, v_m) and $(\bar{v}_1, \bar{v}_2, \dots, \bar{v}_m)$ solve **D**, then there exist constants t_i for $i = 1, 2, \dots, m$ such that $\sum_{i=1}^m t_i = 0$ and $v_i = \bar{v}_i + t_i$, μ_i almost everywhere, for all i .*

Proof. Using the convexification trick in the proof of Theorem 2.1, we can find c -conjugate solutions (u_1, u_2, \dots, u_m) and $(\bar{u}_1, \bar{u}_2, \dots, \bar{u}_m)$ to **D** such that $v_i(x_i) \leq u_i(x_i)$ and $\bar{v}_i(x_i) \leq \bar{u}_i(x_i)$ for all $x_i \in M_i$. Now, as

$$\sum_{i=1}^m \int_{M_i} v_i(x_i) d\mu_i = \sum_{i=1}^m \int_{M_i} u_i(x_i) d\mu_i$$

we must have $v_i = u_i$, μ_i almost everywhere. Similarly, $\overline{v}_i = \overline{u}_i$, μ_i almost everywhere. Now, choose $x_i \in M_i$ where u_i and \overline{u}_i are differentiable. Then there exists x_j for all $j \neq i$ such that

$$(x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_m) \in \text{spt}(\mu);$$

Theorem 2.1 then yields

$$u_i(x_i) - c(x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_m) = - \sum_{j \neq i} u_j(x_j).$$

Because

$$u_i(z_i) - c(x_1, x_2, \dots, x_{i-1}, z_i, x_{i+1}, \dots, x_m) \leq - \sum_{j \neq i} u_j(x_j)$$

for all other $z_i \in M_i$ we must have

$$Du_i(x_i) = D_{x_i} c(x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_m).$$

Similarly,

$$D\overline{u}_i(x_i) = D_{x_i} c(x_1, x_2, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_m),$$

hence $Du_i(x_i) = D\overline{u}_i(x_i)$. As this equality holds for almost all x_i we conclude $u_i(x_i) = \overline{u}_i(x_i) + t_i$ for some constant t_i . Choosing any $(x_1, x_2, \dots, x_m) \in \text{spt}(\mu)$ and noting that

$$\sum_{i=1}^m u_i(x_i) = c(x_1, x_2, \dots, x_m) = \sum_{i=1}^m \overline{u}_i(x_i),$$

we obtain $\sum_{i=1}^m t_i = 0$. □

The next two definitions are straightforward generalizations of concepts borrowed from the two marginal setting.

Definition 2.2. For $i \neq j$, we say that c is (i, j) -twisted if the map $x_j \in M_j \mapsto D_{x_i} c(x_1, x_2, \dots, x_m) \in T_{x_i}^* M_i$ is injective, for all fixed x_k , $k \neq j$.

Definition 2.3. We say that c is (i, j) -non-degenerate if $D_{x_i x_j}^2 c(x_1, x_2, \dots, x_m)$, considered as a map from $T_{x_j} M_j$ to $T_{x_i}^* M_i$, is injective for all (x_1, x_2, \dots, x_m) .

In local coordinates, non-degeneracy simply means that the corresponding matrix of mixed, second order partial derivatives has a non-zero determinant. When this condition holds, the inverse map $T_{x_i}^* M_i \rightarrow T_{x_j} M_j$ will be denoted by $(D_{x_i x_j}^2 c)^{-1}(x_1, x_2, \dots, x_m)$.

When $m = 2$, the non-degeneracy condition is not needed to ensure the existence of an optimal map (although it plays an important role in studying the regularity of that map). On the other hand, the twist condition plays an essential role in showing that Monge's problem has a solution; it ensures that a first order, differential condition arising from the duality theorem can be solved uniquely for one variable as a function of the other [14] (see also [1], [8] and [2]). In light of this, one might expect that, for $m \geq 3$, if c is (i, j) -twisted for all $i \neq j$, then the Kantorovich solution μ induces a Monge solution. This is not true, as our example in [18] demonstrates. In the multi-marginal problem, duality yields m first order conditions; our strategy in this paper is to show that if we fix the first variable, these equations can be uniquely solved for the other $m - 1$ variables. In the problems considered by Gangbo and Świąch [9] and Heinich [10], these equations turn out to have a particularly simple form and can be solved explicitly. For more general cost functions, this becomes a much more subtle issue. Our proof will combine a second order, differential condition with tools from convex analysis and will require that the tensor T , defined below, is negative definite.

Definition 2.4. *Suppose c is $(1, m)$ -non-degenerate. Let $\vec{y} = (y_1, y_2, \dots, y_m) \in M_1 \times M_2 \times \dots \times M_m$. For each $i := 2, 3, \dots, m - 1$ choose a point $\vec{y}(i) = (y_1(i), y_2(i), \dots, y_m(i)) \in \overline{M}_1 \times \overline{M}_2 \times \dots \times \overline{M}_m$ such that $y_i(i) = y(i)$. Define the following bi-linear maps on $T_{y_2} M_2 \times T_{y_3} M_3 \times \dots \times T_{y_{m-1}} M_{m-1}$:*

$$S_{\vec{y}} = - \sum_{j=2}^{m-1} \sum_{\substack{i=2 \\ i \neq j}}^{m-1} D_{x_i x_j}^2 c(\vec{y}) + \sum_{i,j=2}^{m-1} (D_{x_i x_m}^2 c (D_{x_1 x_m}^2 c)^{-1} D_{x_1 x_j}^2 c)(\vec{y})$$

$$H_{\vec{y}, \vec{y}(2), \vec{y}(3), \dots, \vec{y}(m-1)} = \sum_{i=2}^{m-1} (Hess_{x_i} c(\vec{y}(i)) - Hess_{x_i} c(\vec{y}))$$

$$T_{\vec{y}, \vec{y}(2), \vec{y}(3), \dots, \vec{y}(m-1)} = S_{\vec{y}} + H_{\vec{y}, \vec{y}(2), \vec{y}(3), \dots, \vec{y}(m-1)}$$

Note that $D_{x_i x_j}^2 c(x_1, x_2, \dots, x_m)$, $Hess_{x_i} c(x_1, x_2, \dots, x_m)$ and the composition $(D_{x_i x_m} c (D_{x_1 x_m}^2 c)^{-1} D_{x_1 x_j}^2 c)(x_1, x_2, \dots, x_m)$ are actually bi-linear maps on the spaces $T_{x_i} M_i \times T_{x_j} M_j$, $T_{x_i} M_i \times T_{x_i} M_i$ and $T_{x_i} M_i \times T_{x_j} M_j$, respectively, but we can extend them to maps on the product space $(T_{x_2} M_2 \times T_{x_3} M_3 \times \dots \times T_{x_{m-1}} M_{m-1})^2$ by considering only the appropriate components of the tangent vectors.

Though T looks complicated, it appears naturally in our argument. The condition $T < 0$ is in one sense analogous to the twist and non-degeneracy conditions that are so important in the two marginal problem. Like the non-degeneracy condition, negativity of S is an inherently local property on $M_1 \times M_2 \times \dots \times M_m$; under this condition, one can show that our system of equations is locally uniquely solvable. To show that the solution is actually globally unique requires something more; in the two marginal case, this is the twist condition, which can be seen as a global extension of non-degeneracy. In our setting, requiring that the sum $T = S + H < 0$ turns out to be enough to ensure that the locally unique solution is in fact globally unique.

3 Monge solutions

We are now in a position to precisely state our main theorem:

Theorem 3.1. *Suppose that:*

1. *c is $(1, m)$ -non-degenerate.*
2. *c is $(1, m)$ -twisted.*
3. *For all choices of $\vec{y} = (y_1, y_2, \dots, y_m) \in M_1 \times M_2 \times \dots \times M_m$ and of $\vec{y}(i) = (y_1(i), y_2(i), \dots, y_m(i)) \in \overline{M_1} \times \overline{M_2} \times \dots \times \overline{M_m}$ such that $y_i(i) = y_i$ for $i = 2, \dots, m-1$, we have*

$$T_{\vec{y}, \vec{y}(2), \vec{y}(3), \dots, \vec{y}(m-1)} < 0. \quad (3)$$

4. *The first marginal μ_1 does not charge sets of Hausdorff dimension less than or equal to $n-1$.*

Then any solution μ to the Kantorovich problem is concentrated on the graph of a function; that is, there exist functions $G_i : M_1 \rightarrow M_i$ such that $\text{graph}(\vec{G}) = \{(x_1, G_2(x_1), G_3(x_1), \dots, G_m(x_1))\}$ satisfies $\mu(\text{graph}(\vec{G})) = 1$

Proof. Let u_i be a c -conjugate solution to the dual problem. Now, u_1 is semi-concave and hence differentiable off a set of Hausdorff dimension $n - 1$; as μ_1 vanishes on every set of Hausdorff dimension less than or equal to $n - 1$, by Theorem 2.1 it suffices to show that for every $x_1 \in M_1$ where u_1 is differentiable, there is at most one $(x_2, x_3, \dots, x_m) \in M_2 \times M_3 \times \dots \times M_m$ such that $\sum_{i=1}^m u_i(x_i) = c(x_1, x_2, x_3, \dots, x_m)$. Note that this equality implies that $D_{x_i} c(x_1, x_2, \dots, x_m) \in \bar{\partial} u_i(x_i)$ for all $i = 1, 2, \dots, m$; in particular, as u_1 is differentiable at x_1 , $Du_1(x_1) = D_{x_1} c(x_1, x_2, \dots, x_m)$. Our strategy will be to show that these inclusions can hold for at most one (x_2, x_3, \dots, x_m) .

Fix a point x_1 where u_1 is differentiable. Twistedness implies that the equation $Du_1(x_1) = D_{x_1} c(x_1, x_2, \dots, x_m)$ defines x_m as a function $x_m = F_{x_1}(x_2, \dots, x_{m-1})$ of the variables x_2, x_3, \dots, x_{m-1} ; non-degeneracy and the implicit function theorem then imply that F_{x_1} is continuously differentiable with respect to x_2, x_3, \dots, x_{m-1} and

$$D_{x_i} F_{x_1}(x_2, \dots, x_{m-1}) = -((D_{x_1 x_m}^2 c)^{-1} D_{x_1 x_i}^2 c)(x_1, x_2, \dots, F_{x_1}(x_2, \dots, x_{m-1}))$$

for $i = 2, \dots, m - 1$. We will show that there exists at most one point $(x_2, x_3, \dots, x_{m-1}) \in M_2 \times M_3 \times \dots \times M_{m-1}$ such that

$$D_{x_i} c(x_1, x_2, \dots, F_{x_1}(x_2, \dots, x_{m-1})) \in \bar{\partial} u_i(x_i)$$

for all $i = 2, \dots, m - 1$.

The proof is by contradiction; suppose there are two such points, $(x_2, x_3, \dots, x_{m-1})$ and $(\bar{x}_2, \bar{x}_3, \dots, \bar{x}_{m-1})$. For $i = 2, \dots, m - 1$, we can choose Riemannian geodesics $\gamma_i(t)$ in M_i such that $\gamma_i(0) = x_i$ and $\gamma_i(1) = \bar{x}_i$. Take a measurable selection of covectors $V_i(t) \in \partial u_i(\gamma_i(t))$. We will show that $f(1) < f(0)$, where

$$f(t) := \sum_{i=2}^{m-1} [V_i(t) - D_{x_i} c(x_1, \vec{\gamma}(t))] \langle \frac{d\gamma_i}{dt} \rangle$$

and we have used $(x_1, \vec{\gamma}(t))$ as a shorthand for $(x_1, \gamma_2(t), \dots, \gamma_{m-1}(t), F_{x_1}(\gamma_2(t), \dots, \gamma_{m-1}(t)))$ and $\langle a, b \rangle$ to denote the duality pairing between a 1-form a and a vector b . This will clearly imply the desired result.

For each t and each $i = 2, \dots, m - 1$, by c -conjugacy of u_i and the compactness of $\overline{M_j}$ for $j \neq i$, we can choose $\vec{y}(i; t) = (y_1(i; t), y_2(i; t), \dots, y_m(i; t)) \in \overline{M_1} \times \overline{M_2} \times \dots \times \overline{M_m}$ so that $y_i(i; t) = \gamma_i(t)$ and

$$\sum_{j=1}^m u_j(y_j(i; t)) = c(y_1(i; t), y_2(i; t), \dots, y_m(i; t))$$

Note that $V_i(t)\langle \frac{d\gamma_i}{dt} \rangle$ supports the semi-concave function $T \in [0, 1] \mapsto u_i(\gamma_i(t))$. But $u_i(\gamma_i(t))$ is twice differentiable almost everywhere and hence we have $V_i(t)\langle \frac{d\gamma_i}{dt} \rangle = \frac{d(u_i(\gamma_i(t)))}{dt}$ for almost all t and, by semi-concavity, $V_i(1)\langle \frac{d\gamma_i}{dt} \rangle - V_i(0)\langle \frac{d\gamma_i}{dt} \rangle \leq \int_0^1 \frac{d^2(u_i(\gamma_i(t)))}{dt^2} dt$. Now, for any $t, s \in [0, 1]$

$$u_i(\gamma_i(t)) \leq c(y_1(i; s), y_2(i; s), \dots, y_{i-1}(i; s), \gamma_i(t), y_{i+1}(i; s), \dots, y_m(i; s)) - \sum_{j \neq i} u_j(y_j(i; s))$$

and we have equality when $t = s$, as $\gamma_i(s) = y_i(i; s)$. Hence, whenever $\frac{d^2(u_i(\gamma_i(t)))}{dt^2}$ exists, we have

$$\begin{aligned} \left. \frac{d^2(u_i(\gamma_i(t)))}{dt^2} \right|_{t=s} &\leq \left. \frac{d^2(c(y_1(i; s), y_2(i; s), \dots, y_{i-1}(i; s), \gamma_i(t), y_{i+1}(i; s), \dots, y_m(i; s)))}{dt^2} \right|_{t=s} \\ &= Hess_{x_i} c(y_1(i; s), y_2(i; s), \dots, y_m(i; s)) \langle \frac{d\gamma_i}{ds}, \frac{d\gamma_i}{ds} \rangle \end{aligned}$$

We conclude that

$$V_i(1)\langle \frac{d\gamma_i}{dt} \rangle - V_i(0)\langle \frac{d\gamma_i}{dt} \rangle \leq \int_0^1 Hess_{x_i} c(y_1(i; t), y_2(i; t), \dots, y_m(i; t)) \langle \frac{d\gamma_i}{dt}, \frac{d\gamma_i}{dt} \rangle dt \quad (4)$$

Turning now to the other term in $f(1) - f(0)$, we have

$$\begin{aligned} &D_{x_i} c(x_1, \vec{\gamma}(1)) \langle \frac{d\gamma_i}{dt} \rangle - D_{x_i} c(x_1, \vec{\gamma}(0)) \langle \frac{d\gamma_i}{dt} \rangle \\ &= \int_0^1 \frac{d}{dt} (D_{x_i} c(x_1, \vec{\gamma}(t)) \langle \frac{d\gamma_i}{dt} \rangle) dt \\ &= \int_0^1 \left(\sum_{\substack{j=2 \\ j \neq i}}^{m-1} \left(D_{x_i x_j}^2 c(x_1, \vec{\gamma}(t)) \right) \langle \frac{d\gamma_i}{dt}, \frac{d\gamma_j}{dt} \rangle + Hess_{x_i} c(x_1, \vec{\gamma}(t)) \langle \frac{d\gamma_i}{dt}, \frac{d\gamma_i}{dt} \rangle \right. \\ &\quad \left. + \sum_{j=2}^{m-1} \left(D_{x_i x_m}^2 c(x_1, \vec{\gamma}(t)) D_{x_j F_{x_1}}(\vec{\gamma}(t)) \right) \langle \frac{d\gamma_i}{dt}, \frac{d\gamma_j}{dt} \rangle \right) dt \\ &= \int_0^1 \left(\sum_{\substack{j=2 \\ j \neq i}}^{m-1} \left(D_{x_i x_j}^2 c(x_1, \vec{\gamma}(t)) \right) \langle \frac{d\gamma_i}{dt}, \frac{d\gamma_j}{dt} \rangle + Hess_{x_i} c(x_1, \vec{\gamma}(t)) \langle \frac{d\gamma_i}{dt}, \frac{d\gamma_i}{dt} \rangle \right. \\ &\quad \left. - \sum_{j=2}^{m-1} \left((D_{x_i x_m}^2 c(D_{x_1 x_m}^2 c)^{-1} D_{x_1 x_j}^2 c)(x_1, \vec{\gamma}(t)) \right) \langle \frac{d\gamma_i}{dt}, \frac{d\gamma_j}{dt} \rangle \right) dt \quad (5) \end{aligned}$$

Combining (4) and (5) yields

$$\begin{aligned} f(1) - f(0) &\leq \int_0^1 T_{(x_1, \vec{\gamma}(t)), \vec{y}(2;t), \vec{y}(3;t), \dots, \vec{y}(m-1;t)} \left\langle \frac{d\vec{\gamma}}{dt}, \frac{d\vec{\gamma}}{dt} \right\rangle dt \\ &< 0 \end{aligned}$$

□

Corollary 3.2. *Under the same conditions as Theorem 1, the Monge problem \mathbf{M} admits a unique solution and the solution to the Kantorovich problem \mathbf{K} is unique.*

Proof. We first show that the G_i defined in Theorem 3.1 push μ_1 to μ_i for all $i = 2, 3, \dots, m$. Pick a Borel set $B \in M_i$. We have

$$\begin{aligned} \mu_i(B) &= \mu(M_1 \times M_2 \times \dots \times M_{i-1} \times B \times M_{i+1} \times \dots \times M_m) \\ &= \mu\left((M_1 \times M_2 \times \dots \times M_{i-1} \times B \times M_{i+1} \times \dots \times M_m) \cap \text{graph}(\vec{G})\right) \\ &= \mu\left(\{(x_1, G_2(x_1), \dots, G_m(x_1)) \mid G_i(x_1) \in B\}\right) \\ &= \mu\left((G_i^{-1}(B) \times M_2 \times \dots \times M_m) \cap \text{graph}(\vec{G})\right) \\ &= \mu(G_i^{-1}(B) \times M_2 \times \dots \times M_m) \\ &= \mu_1(G_i^{-1}(B)) \end{aligned}$$

This implies that (G_2, G_3, \dots, G_m) solves \mathbf{M} . To prove uniqueness of μ , note that any other optimizer $\bar{\mu}$ must also be concentrated on $\text{graph}(\vec{G})$, which in turn implies $\bar{\mu} = (\text{Id}, G_2, \dots, G_m)_\# \mu_1 = \mu$. Uniqueness of (G_2, G_3, \dots, G_m) now follows immediately; if $(\bar{G}_2, \bar{G}_3, \dots, \bar{G}_m)$ is another solution to \mathbf{M} then $(\text{Id}, \bar{G}_2, \bar{G}_3, \dots, \bar{G}_m)_\# \mu_1$ is another solution to \mathbf{K} , which must then be concentrated on $\text{graph}(\vec{G})$. This means that $G_i = \bar{G}_i$, μ_1 almost everywhere. □

4 Examples

In this section, we discuss several types of cost functions to which Theorem 3.1 applies. In these examples, the complicated tensor T simplifies considerably.

Example 4.1. (*Perturbations of concave functions of the sum*) Gangbo and Świąch [9] and Heinich [10] treated cost functions defined on $(\mathbb{R}^n)^m$ by $c(x_1, x_2, \dots, x_m) = h(\sum_{k=1}^m x_k)$ where $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is strictly concave. Here, we make the slightly stronger assumption that h is C^2 with $D^2h < 0$. Assuming each μ_i is compactly supported, we can take each M_i to be a bounded, convex domain in \mathbb{R}^n . Now, $D_{x_i}c(x_1, x_2, \dots, x_m) = Dh(\sum_{k=1}^m x_k)$ and $D^2_{x_i x_j}c(x_1, x_2, \dots, x_m) = D^2h(\sum_{k=1}^m x_k)$, where we have made the obvious identification between tangent spaces at different points. c is then clearly $(1, m)$ -twisted and $(1, m)$ -non-degenerate. Furthermore, the bi-linear map $S_{\vec{y}}$ on $(\mathbb{R}^n)^{m-2}$ is block diagonal, and each of its diagonal blocks is

$$D^2h(\sum_{k=1}^m y_k).$$

Similarly, as $Hess_{x_i}c(\vec{y}(i)) = D^2h(\sum_{k=1}^m y_k(i))$ and $Hess_{x_i}c(\vec{y}) = D^2h(\sum_{k=1}^m y_k)$, $H_{\vec{y}, \vec{y}(2), \vec{y}(3), \dots, \vec{y}(m-1)}$ is block diagonal and its i th diagonal block is

$$D^2h(\sum_{k=1}^m y_k(i)) - D^2h(\sum_{k=1}^m y_k).$$

Therefore, $T_{\vec{y}, \vec{y}(2), \vec{y}(3), \dots, \vec{y}(m-1)}$ is block diagonal and its i th diagonal block is

$$D^2h(\sum_{k=1}^m y_k(i)).$$

This is clearly negative definite. Furthermore, C^2 perturbations of this cost function will also satisfy $T_{\vec{y}, \vec{y}(2), \vec{y}(3), \dots, \vec{y}(m-1)} < 0$; this shows that the results of Gangbo and Świąch and Heinich are robust with respect to perturbations of the cost function.

Example 4.2. (*Bi-linear costs*) We now turn to bi-linear costs; suppose $c : (\mathbb{R}^n)^m \rightarrow \mathbb{R}$ is given by $c(x_1, x_2, \dots, x_m) = \sum_{i \neq j} (x_i)^T A_{ij} x_j$ for n by n matrices A_{ij} . If A_{1m} is non-singular, c is $(1, m)$ -twisted and $(1, m)$ -non-degenerate. Now, the Hessian terms in T vanish and so the condition $T < 0$ becomes a condition on the A_{ij} . For example, when $m = 3$, we have $T = A_{21}(A_{31})^{-1}A_{32}$; $T < 0$ is the same condition that ensures the solution to \mathbf{K} is contained in an n dimensional submanifold in [18].

Note that after changing coordinates in x_2 and x_3 , we can assume any bi-linear three-marginal cost is of the form

$$c(x_1, x_2, x_3) = x_1 \cdot x_2 + x_1 \cdot x_3 + x_2^T A x_3$$

In these coordinates, the threefold product $A_{21}(A_{31})^{-1}A_{32} = A^T$. Applying the linear change of coordinates

$$\begin{aligned} x_1 &\mapsto U_1 x_1 \\ x_2 &\mapsto U_2 x_2 \\ x_3 &\mapsto U_3 x_3 \end{aligned}$$

yields

$$c(x_1, x_2, x_3) = x_1^T U_1^T U_2 x_2 + x_1^T U_1^T U_3 x_3 + x_2^T U_2^T A U_3 x_3$$

If A is negative definite and symmetric, then we can choose $U_3 = U_2$ such that $U_2^T A U_3 = -I$ and $U_1 = -(U_2^T)^{-1}$ to obtain

$$c(x_1, x_2, x_3) = -x_1^T x_2 - x_1^T x_3 - x_2^T x_3$$

which is equivalent² to the cost of Gangbo and Świąch. As the symmetry of $D_{x_2 x_1}^2 c (D_{x_3 x_1}^2 c)^{-1} D_{x_3 x_2}^2 c$ is independent of our choice of coordinates, we conclude that c is equivalent to Gangbo and Świąch's cost if and only if $A_{21}(A_{31})^{-1}A_{32}$ is symmetric and negative definite. Thus, when $m = 3$ our result restricted to bi-linear costs generalizes Gangbo and Świąch's theorem from costs for which $A_{21}(A_{31})^{-1}A_{32}$ is symmetric and negative definite to ones for which it is only negative definite.

Example 4.3. There is another class of three marginal problems which Theorem 3.1 applies to: on $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$, set

$$c(x_1, x_2, x_3) = g(x_1, x_3) + \frac{|x_1 - x_2|^2}{2} + \frac{|x_3 - x_2|^2}{2}.$$

If $g(x_1, x_3) = \frac{|x_1 - x_3|^2}{2}$, this is equivalent to the cost of Gangbo and Świąch. More generally, if g is $(1, 3)$ -twisted and non-degenerate, then c is as well.

²We say cost functions c and \bar{c} are equivalent if $\bar{c}(x_1, x_2, \dots, x_m) = c(x_1, x_2, \dots, x_m) + \sum_{i=1}^m g_i(x_i)$. As the effect of the g_i 's is to shift the functionals $C(G_2, G_3, \dots, G_m)$ and $C(\mu)$ by the constant $\sum_{i=1}^m \int_{M_i} g_i(x_i) d\mu_i$, studying c is essentially equivalent to studying \bar{c} .

Moreover, if we make the usual identification between tangent spaces at different points in \mathbb{R}^n , we have

$$T_{\vec{y}, \vec{y}(2)} = (D_{x_1 x_3}^2 g(y_1, y_3))^{-1}.$$

Hence, if $D_{x_1 x_3}^2 g(y_1, y_3) < 0$, we have $T_{\vec{y}, \vec{y}(2)} < 0$. This will be the case if, for example, $g(x_1, x_3) = h(x_1 - x_3)$ for h uniformly convex or $g(x_1, x_3) = h(x_1 + x_3)$ for h uniformly concave.

Example 4.4. (*Hedonic Pricing*) Chiappori, McCann and Nesheim [6] and Carlier and Ekeland [4] showed that finding equilibrium in a certain hedonic pricing model is equivalent to solving a multi-marginal optimal transportation problem with a cost function of the form $c(x_1, x_2, \dots, x_m) = \inf_{z \in Z} \sum_{i=1}^m f_i(x_i, z)$. Let us assume:

1. Z is a C^2 smooth n -dimensional manifold.
2. For all i , f_i is C^2 and non-degenerate.
3. For each (x_1, x_2, \dots, x_m) the infimum is attained by a unique $z(x_1, x_2, \dots, x_m) \in Z$ and
4. $\sum_{i=1}^m D_{zz}^2 f_i(x_i, z(x_1, x_2, \dots, x_m))$ is non-singular.

In [18], we showed that these conditions implied that c is C^2 and (i, j) -non-degenerate for all $i \neq j$; we then showed that the support of any optimizer is contained in an n -dimensional Lipschitz submanifold of the product $M_1 \times M_2 \times \dots \times M_m$. Here we examine conditions on the f_i that ensure the hypotheses of Theorem 3.1 are satisfied. If, for fixed $i \neq j$, we assume in addition that:

5. f_i is x_i, z twisted (that is, $z \mapsto D_{x_i} f_i(x_i, z)$ is injective) and
6. f_j is z, x_j twisted.

then c is (i, j) -twisted. Indeed, note that $c(x_1, x_2, \dots, x_m) \leq \sum_{i=1}^m f_i(x_i, z)$ with equality when $z = z(x_1, x_2, \dots, x_m)$; therefore,

$$D_{x_i} c(x_1, x_2, \dots, x_m) = D_{x_i} f(x_i, z(x_1, x_2, \dots, x_m)) \quad (6)$$

Therefore, for fixed x_k for all $k \neq j$, the map $x_j \mapsto D_{x_i} c(x_1, x_2, \dots, x_m)$ is the composition of the maps $x_j \mapsto z(x_1, x_2, \dots, x_m)$ and $z \mapsto D_{x_i} f(x_i, z)$. The later map is injective by assumption. Now, note that

$$\sum_{k=1}^m D_z f_i(x_i, z(x_1, x_2, \dots, x_m)) = 0;$$

hence,

$$D_z f_j(x_j, z(x_1, x_2, \dots, x_m)) = - \sum_{k \neq j} D_z f_k(x_k, z(x_1, x_2, \dots, x_m)).$$

Twistedness of f_j now immediately implies injectivity of the first map.

We now investigate the form of the tensor T .

As $A(x_1, x_2, \dots, x_m) := \sum_{i=1}^m D_{zz}^2 f_i(x_i, z(x_1, x_2, \dots, x_m))$ is non-singular by assumption, the implicit function theorem implies that $z(x_1, x_2, \dots, x_m)$ is differentiable and

$$D_{x_i} z(x_1, x_2, \dots, x_m) = - (A(x_1, x_2, \dots, x_m))^{-1} D_{zx_i}^2 f_i(x_i, z(x_1, x_2, \dots, x_m))$$

Furthermore, note that as A is positive semi-definite by the minimality of $z \mapsto \sum_{i=1}^m f_i(x_i, z)$ at $z(x_1, x_2, \dots, x_m)$, the non-singular assumption implies that it is in fact positive definite.

Differentiating (6) with respect to x_i for $i = 2, 3, \dots, m-1$ yields:

$$\text{Hess}_{x_i} c = -(D_{x_i z}^2 f_i) D_{x_i} z + \text{Hess}_{x_i} f_i = -(D_{x_i z}^2 f_i) A^{-1} (D_{zx_i}^2 f_i) + \text{Hess}_{x_i} f_i.$$

where we have suppressed the arguments x_1, x_2, \dots, x_m and $z(x_1, x_2, \dots, x_m)$. A similar calculation yields, for all $i \neq j$,

$$D_{x_i x_j}^2 c = (D_{x_i z}^2 f_i) D_{x_j} z = -(D_{x_i z}^2 f_i) A^{-1} (D_{zx_j}^2 f_j)$$

Thus, for all $i \neq j$, a straightforward calculation yields

$$D_{x_i x_m}^2 c (D_{x_1 x_m}^2 c)^{-1} D_{x_1 x_j}^2 c = -(D_{x_i z}^2 f_i) A^{-1} (D_{zx_j}^2 f_j) = D_{x_i x_j}^2 c,$$

Hence, $S_{\vec{y}}$ is block diagonal. Furthermore, another simple calculation implies that its i th diagonal block is

$$\left[D_{x_i x_m}^2 c (D_{x_1 x_m}^2 c)^{-1} D_{x_1 x_i}^2 c \right] (\vec{y}) = - \left[(D_{x_i z}^2 f_i) A^{-1} (D_{zx_i}^2 f_i) \right] (\vec{y}, z(\vec{y})).$$

In addition, $H_{\vec{y}, \vec{y}(2), \vec{y}(3), \dots, \vec{y}(m-1)}$ is block diagonal and its i th block is

$$\begin{aligned} & - \left[(D_{x_i z}^2 f_i) A^{-1} (D_{zx_i}^2 f_i) \right] (\vec{y}(i), z(\vec{y}(i))) + \text{Hess}_{x_i} f_i(y_i, z(\vec{y}(i))) \\ & + \left[(D_{x_i z}^2 f_i) A^{-1} (D_{zx_i}^2 f_i) \right] (\vec{y}, z(\vec{y})) - \text{Hess}_{x_i} f_i(y_i, z(\vec{y})) \end{aligned}$$

Hence, $T_{\vec{y}, \vec{y}(2), \vec{y}(3), \dots, \vec{y}(m-1)}$ is block diagonal and its i th block is

$$-\left[(D_{x_i z} f_i)A^{-1}(D_{z x_i}^2 f_i)\right]\left(\vec{y}(i), z(\vec{y}(i))\right) + \text{Hess}_{x_i} f_i\left(y_i, z(\vec{y}(i))\right) - \text{Hess}_{x_i} f_i\left(y_i, z(\vec{y})\right) \quad (7)$$

Therefore, $T_{\vec{y}, \vec{y}(2), \vec{y}(3), \dots, \vec{y}(m-1)}$ is negative definite if and only if each of its diagonal blocks is. Now, A is symmetric and positive definite; therefore A^{-1} is as well. The first term in the i th block of (7) is therefore negative definite; the entire block will be negative definite if this term dominates the difference of the Hessian terms. This is the case if, for example, $M_i = \mathbb{R}^n$ and f_i takes the form $f_i(x_i, z) = x_i \alpha_i(z) + \beta_i(x_i) + \lambda_i(z)$ for all $i = 2, 3, \dots, m-1$, in which case $\text{Hess}_{x_i} f_i\left(y_i, z(\vec{y}(i))\right) = \text{Hess}_{x_i} f_i\left(y_i, z(\vec{y})\right)$.

References

- [1] Brenier, Y. Decomposition polaire et rearrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris. Ser. I Math.*, **305** (1987), 805-808.
- [2] Caffarelli, L., Allocation maps with general cost functions. In *Partial Differential Equations and Applications* (P. Marcellini, G. Talenti and E. Vesintini, eds.), Lecture Notes in Pure and Applied Math. 177 (1996), pp. 29-35.
- [3] Carlier, G., *On a class of multidimensional optimal transportation problems*, J. Convex Anal. 10, 2003, no. 2, pp. 517-529.
- [4] Carlier, G., and Ekeland, I., *Matching for teams*, Econom. Theory. 42, 2010, no. 2, pp. 397-418.
- [5] Carlier, G., and Nazaret, B., *Optimal transportation for the determinant* ESAIM Control Optim. Calc. Var. 14, 2008, no.4, pp. 678-698.
- [6] Chiappori, P-A., McCann, R., and Nesheim, L., *Hedonic price equilibria, stable matching and optimal transport; equivalence, topology and uniqueness*, Econom. Theory. 42, 2010, no. 2, pp. 317-354.
- [7] Gangbo, W. *Habilitation Thesis*, Universite de Metz, (1995).

- [8] Gangbo, W. and McCann, R.J. The geometry of optimal transportation. *Acta Math.*, **177** (1996), 113-161.
- [9] Gangbo, W., and Świąch, A., *Optimal maps for the multidimensional Monge-Kantorovich problem*, Comm. Pure Appl. Math. 51, 1998, no.1, pp. 23-45.
- [10] Heinich, H., *Probleme de Monge pour n probabilités*, C.R. Math. Acad. Sci. Paris 334, 2002, no. 9, pp. 793-795.
- [11] Kellerer, H.G., *Duality theorems for marginal problems*. Z. Wahrsch. Verw. Gebiete 67, 1984, pp. 399-432.
- [12] Kim, Y-H., McCann, R., *Continuity, curvature, and the general covariance of optimal transportation*, to appear in J. Eur. Math. Soc.
- [13] Knott, M., and Smith, C. *On a generalization of cyclic monotonicity and distances among random vectors*. Linear Algebra Appl. 199, 1994, pp. 363-371.
- [14] Levin, V. *Abstract cyclical monotonicity and Monge solutions for the general Monge-Kantorovich problem*. Set-Valued Analysis, 7, no.1, 1999, pp. 7-32.
- [15] Ma, X-N., Trudinger, N., and Wang, X-J. *Regularity of potential functions of the optimal transportation problem*. Arch. Rational Mech. Anal., 177 (2005), 151-183.
- [16] McCann, R. *Polar factorization of maps on Riemannian manifolds*. Geom. Funct. Anal. 11, 2001, pp. 589-608.
- [17] Olkin, I., and Rachev, S.T., *Maximum submatrix traces for positive definite matrices*. SIAM J. Matrix Ana. Appl. 14, 1993, pp. 390-397.
- [18] Pass, B. *On the local structure of optimal measures in the multi-marginal optimal transportation problem*. Preprint available at arXiv:1005.2162.
- [19] Rüschendorf, L. and Uckelmann, L. *On optimal multivariate couplings*. In *Proceedings of Prague 1996 conference on marginal problems* (eds. V. Benes, I. Stepan) Kluwer Acad. Publ. (1997), pp. 261-274.

- [20] Rüschendorf, L. and Uckelmann, L., *On the n -coupling problem*. Journal of Multivariate Analysis 81, 2002, pp. 242-258.
- [21] Villani, C., *Topics in optimal transportation*, volume 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence, 2003.
- [22] Villani, C., *Optimal transport: old and new*, volume 338 of Grundlehren der mathematischen Wissenschaften. Springer, New York, 2009.